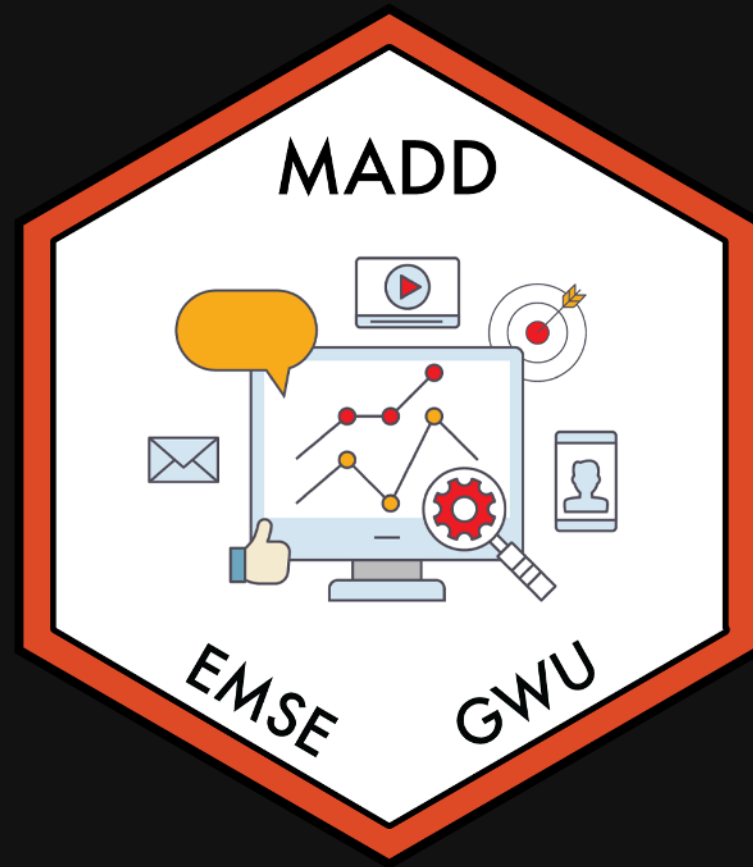




Utility Models

🏛️ EMSE 6035: Marketing Analytics for Design Decisions

👤 John Paul Helveston

📅 October 14, 2026

Week 8: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters
4. No choice
5. Team project utility models

Week 8: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters
4. No choice
5. Team project utility models

Random utility model

The utility for alternative j is

$$\tilde{u}_j = v_j + \varepsilon_j$$

v_j = Things we observe (non-random variables)

ε_j = Things we *don't* observe (random variable)

$$\mathcal{U}_{\tilde{j}} = \mathcal{U}_j + \varepsilon \tilde{j}$$

Practice Question 1

- a. A random variable, $x\sim$, has the PDF, $f_x(x)$. Write the equation to compute its total probability (hint: think area under the curve!). What is the answer to the equation?
- b. A random variable, $x\sim$, has a uniform distribution between the values 0 and 1. Draw the probability density function (PDF) and Cumulative Density Function (CDF) of $x\sim$.
- c. The value of a random variable, $x\sim$, is determined by rolling one fair, 6-sided dice. Draw the PDF and CDF of $x\sim$.

Logit model: Assume that $\varepsilon_j \sim$ Gumbel Distribution

Practice Question 2

a) A consumer is making a choice between two bars of chocolate:

- Milk chocolate (m)
- Dark chocolate (d)

Assume that we know the observed utility of each bar to be $v_m = 3$ and $v_d = 4$. Using a logit model, compute the probabilities of choosing each bar: P_m and P_d .

b) A third bar of chocolate is now added to the choice set. It is the exact same as the milk chocolate bar, but it has a slightly different wrapper (which has no effect on the consumer's utility). Now, $v_{m1} = v_{m2} = 3$, and $v_d = 4$. Based on the probabilities from question a), what would we expect the probabilities of choosing each bar to be? What probabilities does the logit model produce?

“Observed utility” (v_j) **is a weighted sum of attribute values**

$$v_j = \beta_1 x_j^A + \beta_2 x_j^B + \dots$$

Each x_j is an observable attribute (*price*, etc.)

We know $x_j^A, x_j^B, \dots$,
we want to *estimate* $\beta_1, \beta_2, \dots$

Notation Convention

Continuous: x_j

$$u_j = \beta_1 x_j^{\text{price}} + \dots$$

Discrete: δ_j

$$u_j = \beta_1 \delta_j^{\text{ford}} + \beta_2 \delta_j^{\text{gm}} \dots$$

```
#>   price
#> 1     1
#> 2     2
#> 3     3
```

```
#>   brand brand_BMW brand_Ford brand_GM
#> 1  Ford         0         1         0
#> 2   GM         0         0         1
#> 3  BMW         1         0         0
```

Practice Question 3

Attribute	Bar 1	Bar 2	Bar 3
Price	\$1.20	\$1.50	\$3.00
% Cacao	10%	60%	80%

- Write out a model for the *observed* utility of each chocolate bar in the above set.
- If the coefficient for the *price* attribute was -0.1 and the coefficient for % *Cacao* attribute was 0.1, what is the difference in the observed utility between bars 3 and 1?
- With the addition of the *brand* attribute, repeat part a.

Attribute	Bar 1	Bar 2	Bar 3
Price	\$1.20	\$1.50	\$3.00
% Cacao	10%	60%	80%
Brand	Hershey	Lindt	Ghirardelli

Your Turn

20:00

Let's say our utility function is:

$$v_j = \beta_1 x_j^{\text{price}} + \beta_2 x_j^{\text{cacao}} + \beta_3 \delta_j^{\text{hershey}} + \beta_4 \delta_j^{\text{lindt}}$$

And we estimate the following coefficients:

Parameter	Coefficient
β_1	-0.1
β_2	0.1
β_3	-2.0
β_4	-0.1

- a. What are the expected probabilities of choosing each of these bars using a logit model?

Attribute	Bar 1	Bar 2	Bar 3
Price	\$1.20	\$1.50	\$3.00
% Cacao	10%	60%	80%
Brand	Hershey	Lindt	Ghirardelli

- b. What price would Bar 2 have to be to get a 50% market share?

Week 8: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters
4. No choice
5. Team project utility models

Download the [logitr-cars](#) repo from GitHub

Exploring choice data

1. Open the `logitr-cars` folder in Positron
2. Open `code/2.1-explore-data.R`

Week 8: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters
4. No choice
5. Team project utility models

Dummy-coded variables

Dummy coding: 1 = "Yes", 0 = "No"

Data frame with one variable: *price*

```
1 data <- data.frame(  
2   price = c(10, 20, 30)  
3 )  
4  
5 data
```

```
#>   price  
#> 1    10  
#> 2    20  
#> 3    30
```

Add dummy columns for each price "level"

```
1 library(fastDummies)  
2  
3 dummy_cols(data, "price")
```

```
#>   price price_10 price_20 price_30  
#> 1    10         1         0         0  
#> 2    20         0         1         0  
#> 3    30         0         0         1
```

Model *price* as continuous

$$v_j = \beta_1 x^{\text{price}}$$

```
1 model <- logitr(  
2   data = data,  
3   choice = "choice",  
4   obsID = "obsID",  
5   pars = "price"  
6 )
```

Coef.	Interpretation
-------	----------------

β_1	how utility changes with increasing <i>price</i>
-----------	--

Model *price* as discrete

$$v_j = \beta_1 \delta^{\text{price}=20} + \beta_2 \delta^{\text{price}=30}$$

```
1 model <- logitr(  
2   data = data,  
3   choice = "choice",  
4   obsID = "obsID",  
5   pars = c("price_20", "price_30")  
6 )
```

Reference level: *price=10*

Coef.	Interpretation
-------	----------------

β_1	utility for <i>price=20</i> relative to <i>price=10</i>
-----------	---

β_2	utility for <i>price=30</i> relative to <i>price=10</i>
-----------	---

Dummy coding categorical variables with `cbc_encode()`

Standard coding (human-readable)

```
1 design |>
2   select(altID, powertrain)
```

```
#>   altID powertrain
#> 1     1   Gasoline
#> 2     2    Hybrid
#> 3     3   Electric
```

Dummy coding (what `logitr()` needs)

```
1 design |>
2   cbc_encode(coding = "dummy") |>
3   select(altID, starts_with("powertrain"))
```

```
#>   altID powertrainHybrid powertrainElectric
#> 1     1             0             0
#> 2     2             1             0
#> 3     3             0             1
```

Rule:

Use `dummy_cols()` for any data frame (simulated or from a file)

Use `cbc_encode()` to preview {cbcTools} designs, e.g. with a *no choice* option

Choosing the reference level

Default: first level is the reference

```
1 cbc_encode(  
2     design,  
3     coding = "dummy"  
4 )
```

$$v_j = \beta_1 \delta^{\text{hybrid}} + \beta_2 \delta^{\text{electric}}$$

Coef.	Interpretation
-------	----------------

β_1	utility of <i>Hybrid</i> relative to <i>Gasoline</i>
-----------	--

β_2	utility of <i>Electric</i> relative to <i>Gasoline</i>
-----------	--

Pick a different reference level:

```
1 cbc_encode(  
2     design,  
3     coding = "dummy",  
4     ref_levels = list(powertrain = "Electric")  
5 )
```

$$v_j = \beta_1 \delta^{\text{gasoline}} + \beta_2 \delta^{\text{hybrid}}$$

Coef.	Interpretation
-------	----------------

β_1	utility of <i>Gasoline</i> relative to <i>Electric</i>
-----------	--

β_2	utility of <i>Hybrid</i> relative to <i>Electric</i>
-----------	--

Estimating utility models

1. Open the `logitr-cars` folder in Positron
2. Open `code/3.1-model-mnl.R`

Your Turn

20:00

1. Run the code chunk to read in the `data.csv` file in the "data" folder, which contains choice observations from chocolate bars with the following attributes:

Attribute	Description
<code>price</code>	Price in \$
<code>percent_cacao</code>	% Cacao (how "dark" the chocolate is)
<code>crispy_rice</code>	0 or 1 for if the bar contains crispy rice
<code>brand</code>	"Hershey", "Lindt", or "Ghirardelli"

2. Write code to estimate the following utility model:

$$u_j = \beta_1 x_j^{\text{price}} + \beta_2 x_j^{\% \text{cacao}} + \beta_3 \delta_j^{\text{crispy}} + \beta_4 \delta_j^{\text{hershey}} + \beta_5 \delta_j^{\text{lindt}} + \varepsilon_j$$

3. Write code to plot the change in utility for the *price* attribute.

Break

05 : 00

Week 8: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters
4. No choice
5. Team project utility models

Estimating utility models with a *No Choice* option

1. Open the `logitr-cars` folder in Positron
2. Open `code/4.1-model-no_choice.R`

Dummy coding with a *No Choice* option

The “no choice” rows have **NA** for every attribute

```
1 data <- data %>%  
2   dummy_cols("powertrain")
```

powertrain_NA = 1 on the no-choice rows.
Do not include it in the model!

Replace the **NA**s with 0 for the **named** columns only

```
1 data <- data %>%  
2   mutate(  
3     across(  
4       c(  
5         price,  
6         fuelEconomy,  
7         accelTime,  
8         powertrain_Gasoline,  
9         powertrain_Hybrid,  
10        powertrain_Electric  
11      ),  
12     ~ replace_na(.x, 0)  
13   )  
14 )
```

Week 8: *Utility Models*

1. Utility models
2. Exploring choice data
3. Linear & discrete parameters
4. No choice
5. Team project utility models

Simulating choice data

Random choices

```
1 data <- cbc_choices(  
2   design = design  
3 )
```

Choices according to assumed model

$$v_j = -0.7x_j^{\text{price}} + 0.1x_j^{\text{fuelEconomy}} - 0.2x_j^{\text{accelTime}} + 0.5\delta_j^{\text{hybrid}} - 4\delta_j^{\text{electric}}$$

```
1 priors <- cbc_priors(  
2   profiles = profiles,  
3   price = -0.7,  
4   fuelEconomy = 0.1,  
5   accelTime = -0.2,  
6   powertrain = c(  
7     Hybrid = 0.5,  
8     Electric = -4.0  
9   )  
10 )  
11 data <- cbc_choices(  
12   design = design,  
13   priors = priors  
14 )
```

Not sure what coefficients to use?

`cbc_suggest_priors()` proposes a starting point based on your attribute levels

```
1 cbc_suggest_priors(profiles)

1 priors <- cbc_priors(
2   profiles = profiles,
3   price = -0.065,
4   fuelEconomy = 0.065,
5   accelTime = -0.325,
6   powertrain = c(
7     Hybrid = 0.2,
8     Electric = -0.4
9   )
10 )
```

These are only a starting point:

- Continuous attributes get small, same-sized effects
- Categorical levels are relative to the reference level
- **Signs are all positive**, so flip them where you expect a *negative* effect (e.g. price)
- Adjust using your own domain knowledge

Estimate a choice model

$$v_j = \beta_1 x_j^{\text{price}} + \beta_2 x_j^{\text{fuelEconomy}} + \beta_3 x_j^{\text{accelTime}} + \beta_4 \delta_j^{\text{hybrid}} + \beta_5 \delta_j^{\text{electric}}$$

```
1 # Simulated choices use standard coding, so dummy code first
2 library(fastDummies)
3 data <- dummy_cols(data, "powertrain")
4
5 model <- logitr(
6   data = data,
7   outcome = "choice",
8   obsID = "obsID",
9   pars = c(
10     "price",
11     "fuelEconomy",
12     "accelTime",
13     "powertrain_Hybrid",
14     "powertrain_Electric"
15   )
16 )
```

Your Turn

20:00

As a team:

1. Go back to your code from last week where you created your choice questions.
2. Write out a utility model for your project.
3. Write code to simulate data according to your utility model - pick some parameter values (see `simulate-choices.R`).
4. Write code to estimate a model using your simulated data.